

Abstract of Chinese Patent No. 99100625 Published February 12, 2003

Title: Method and System for Inputting Chinese Characters into Computers

This invention provides a method to input Chinese characters into computers. The system includes a regular keyboard or a software-defined keyboard, a Chinese character information processor and a display unit. The input method uses stroke codes formed according to the stroke sequence of a Chinese character. This invention features in the following processes: an intelligent stroke key which can replace any stroke code in the operation of inputting stroke codes; as each stroke code is being entered, the display unit updates and exhibits the corresponding candidate characters and candidate radicals to be used which are arranged according to a pre-determined usage frequency; based on the occurrence frequency of actually inputted Chinese characters and phrases, adjust the usage frequency ordering of Chinese characters and phrases corresponding to the stroke code actually inputted and adjust the display of the usage frequency ordering of the corresponding candidate radicals. After inputting a Chinese character, the remaining candidate Chinese characters are divided into two groups, wherein the first group has candidate characters in the corresponding related phrases following the inputted character and the second group has candidate characters unrelated to the inputted character, and wherein the two groups are separated by a color zone so that any candidate character, either phrase-related candidate characters or unrelated candidate characters, can be inputted without a blank space.

[19] 中华人民共和国国家知识产权局

[51] Int. Cl⁷

G06F 3/00



[12] 发明专利说明书

[21] ZL 专利号 99100625.9

[43] 授权公告日 2003 年 2 月 12 日

[11] 授权公告号 CN 1101567C

[22] 申请日 1999.2.9 [21] 申请号 99100625.9

[30] 优先权

[32] 1998. 2. 13 [33] US [31] 09/023259

[71] 专利权人 字原(加拿大)有限公司

地址 加拿大阿尔伯达省

[72] 发明人 华立驰 潘贤麟 邱卫根

审查员 张静海

[74] 专利代理机构 中科专利商标代理有限责任公司

代理人 张祥龄

权利要求书 7 页 说明书 19 页 附图 10 页

[54] 发明名称 汉字字符输入的方法和系统

[57] 摘要

一种计算机汉字输入方法,利用普通或虚拟键盘及汉字信息处理和显示单元,依汉字笔序利用笔划码元输入汉字,其特征包括以下处理方法:在笔划码元输入操作中代替任意笔划码元的智能笔划键;随笔划码元的输入,在显示单元上逐次更新显示预测引导性的依使用频度排序的相应候选汉字及待用汉字部件;随实际输入汉字、词的频度而随机改变相应于输入笔划码元的汉字、词频度排序及相应汉字部件频度排序的随机频度排序显示;输入一个候选汉字后,余下的候选字分为两组显示,一组是相应该输入汉字的相关词组的下汉字,一组是与在先候选字相同的汉字,并用色标区分开,不用空格即可输入联想及非联想汉字的任意候选汉字。

ISSN 1008-4274

1、一种计算机汉字输入系统，其中包括：

a、依汉字笔序利用笔划码元输入汉字的键盘单元或虚拟键盘单元，所述键盘单元具有一智能键，该智能键能在待输入的汉字中代替任意笔划；

b、汉字信息处理单元；及

5 c、汉字及其它相关信息的显示单元；其特征在于还包括：

(1) 所述智能键还能在汉字部件中代替任意笔划码元的处理单元；

(2) 随着笔画的输入，屏幕上显示出包含有对应于输入笔序笔划的汉字、词和汉字部件，这些汉字、词和汉字部件的显示排序是综合操作员使用这些汉字、词和汉字部件的累积频率以及预先设定的频率而排序的处理单元；

10 (3) 当输入一个候选汉字之后，余下的候选汉字被分为两组显示，第一组是相应该输入汉字的相关词组的下一个汉字，第二组是非联想的候选汉字，以便输入联想汉字或非联想的任意候选汉字的处理单元。

2、根据权利要求1的汉字输入系统，其特征在于所述显示、处理单元还包括用于延续搜索下一组候选汉字和部件的延伸键处理单元，和利用所显示的候选汉字及部件之后的点数表示待搜索候选汉字及部件组数的处理单元。

3、根据权利要求1的汉字输入系统，其特征在于所述显示单元还包括可调整候选汉字和部件的显示个数，和候选汉字和部件从左至右排序显示或从中间到左、右两边排列显示的排列形式的显示处理单元。

4、根据权利要求1的汉字输入系统，其特征在于所述键盘是以定位点为标志，在键位上有序排列输入笔划的。

5、根据权利要求1的汉字输入系统，其特征在于所述虚拟键盘的键位上可

对应显示候选汉字和部件，以便随虚拟键位输入候选汉字。

6、根据权利要求1的汉字输入系统，其特征在于该系统可以输出任意所需的汉字外部标准编码。

7、根据权利要求1的汉字输入系统，其特征在于还包括在选择了一个候选部件后，可显示以所选部件为第一书写结构的复杂部件的处理单元。

8、根据权利要求1的汉字输入系统，其特征在于还包括可用色标区分联想词组和非联想词组，用另外色标区分简单部件和复杂部件的处理单元。

9、根据权利要求1至8之一的汉字输入系统，其特征该系统在于可应用于普通计算机输入键盘和显示系统。

10、根据权利要求1至8之一的汉字输入系统，其特征在于该系统可应用于普通手机或普通电话拨号盘和相应的点阵显示单元。

11、根据权利要求1至8之一的汉字输入系统，其特征在于所述显示单元可应用于其它任何汉字输入系统中。

12、一种计算机汉字输入系统，其中包括：

a、依汉字笔序利用笔划码元或用罗马字母或注音符号代替笔划码元输入汉字的键盘单元或虚拟键盘单元；

b、汉字信息处理单元；及

c、汉字及其它相关信息的显示单元；其特征在于还包括：

(1) 随着音标和笔划的输入，屏幕上显示出包含有对应于输入音标和笔序笔划的汉字、词和汉字部件，这些汉字、词和汉字部件的显示排序是综合操作员使用这些汉字、词和汉字部件的累积频率以及预先设定的频率而排序的处理单元；

(2) 当输入一个候选汉字之后, 余下的候选汉字被分为两组显示, 第一组是相应该输入汉字的相关词组的下一个汉字, 第二组是非联想的侯选汉字, 以便输入联想汉字或非联想汉字的任意侯选汉字的处理单元。

(3) 可用普通话或广东话的音标显示侯选汉字的显示处理单元。

5 13、根据权利要求 12 的汉字输入系统, 其特征在于所述显示、处理单元还包括用于延续搜索下一组侯选汉字和部件的延伸键处理单元, 和利用所显示的侯选汉字及部件之后的点数表示待搜索侯选汉字及部件组数的显示处理单元。

14、根据权利要求 12 的汉字输入系统, 其特征在于所述显示单元还包括可调整侯选汉字和部件的显示个数, 和侯选汉字和部件从左至右排序显示或从中间
10 到左、右两边排列显示的排列形式的显示处理单元。

15、根据权利要求 12 的汉字输入系统, 其特征在于所述键盘是以定位点为标志, 在键位上有序排列输入笔划的。

16、根据权利要求 12 的汉字输入系统, 其特征在于所述虚拟键盘的键位上可对应显示侯选汉字和部件, 以便随虚拟键位输入。

15 17、根据权利要求 12 的汉字输入系统, 其特征在于该系统可以输出任意所需的汉字外部标准编码。

18. 根据权利要求 12 的汉字输入系统, 其特征在于还包括在选择了一个侯选部件后, 可显示以所选部件为第一书写结构的复杂部件的处理单元。

19、根据权利要求 12 的汉字输入系统, 其特征在于还包括可用色标区分联想词组和非联想词组, 用另外色标区分简单部件和复杂部件的处理单元。
20

20、根据权利要求 12 至 19 之一的汉字输入系统, 其特征在于该系统可应用于普通计算机输入键盘和显示系统。

21、根据权利要求 12 至 19 之一的汉字输入系统，其特征在于该系统可应用于普通手机或普通电话拨号盘和相应的点阵显示单元。

22、根据权利要求 12 至 19 之一的汉字输入系统，其特征在于所述显示单元可应用于其它任何汉字输入系统中。

5 23、一种计算机汉字输入方法，该方法利用普通键盘单元或虚拟键盘单元，并利用汉字信息处理单元及显示单元，依汉字笔序利用笔划码元输入汉字，所述键盘单元具有一智能键，该智能键能在待输入的汉字中代替任意笔划，其特征在于还包括：

(1) 所述智能键还能在汉字部件中代替任意笔划码元的智能键处理方法；

10 (2) 随着笔划的输入，屏幕上显示出包含有对应于输入笔序笔划的汉字、词和汉字部件，这些汉字、词和汉字部件的显示排序是综合操作员使用这些汉字、词和汉字部件的累积频率以及预先设定的频率而排序的处理方法；

(3) 当输入一个候选汉字之后，余下的候选汉字被分为两组显示，第一组是相应该输入汉字的相关词组的下一个汉字，第二组是非联想的侯选汉字，以便
15 输入联想汉字或非联想汉字的任意侯选汉字的处理方法。

24、根据权利要求 23 的汉字输入方法，其特征在于所述显示、处理方法还包括用于延续搜索下一组侯选汉字和部件的延伸键处理方法，和利用所显示的侯选汉字及部件之后的点数表示待搜索侯选汉字及汉字部件组数的处理方法。

25、根据权利要求 23 的汉字输入方法，其特征在于所述显示方法还包括可
20 调整侯选汉字和部件的显示个数，和侯选汉字和部件从左至右排序显示或从中间到左、右两边排列显示的排列形式的显示处理方法。

26、根据权利要求 23 的汉字输入方法，其特征在于所述键盘是以定位点为

标志，在键位上有序排列输入笔划的。

27、根据权利要求 23 的汉字输入方法，其特征在于还包括所述虚拟键盘的键位上可对应显示侯选汉字和部件，以便随虚拟键位输入的方法。

28、根据权利要求 23 的汉字输入方法，其特征在于所述信息处理单元可以
5 输出任意所需的汉字外部标准编码。

29、根据权利要求 23 的汉字输入方法，其特征在于在选择了一个侯选部件后，可显示以所选部件为第一书写结构的复杂部件。

30、根据权利要求 23 的汉字输入方法，其特征是，可用色标区分联想词组和非联想词组，用另外色标区分简单部件和复杂部件。

10 31、根据权利要求 23 至 30 之一的汉字输入方法，其特征在于该方法可应用于普通计算机输入键盘和显示方法。

32、根据权利要求 23 至 30 之一的汉字输入方法，其特征在于该方法可应用于普通手机和普通电话拨号盘和相应的点阵显示方法。

15 33、根据权利要求 23 至 30 之一的汉字输入方法，其特征在于所述显示单元可应用于其它任何汉字输入方法中。

34、一种计算机汉字输入方法，该方法利用键盘单元或虚拟键盘单元，并利用汉字信息处理单元及显示单元依汉字笔序利用笔划码元或用罗马字母或注音符代替笔划码元输入汉字；其特征在于还包括：

20 (1) 随着音标和笔划的输入，屏幕上显示出包含有对应于输入音标和笔序笔划的汉字、词和汉字部件，这些汉字、词和汉字部件的显示排序是综合操作员使用这些汉字、词和汉字部件的累积频率以及预先设定的频率而排序的处理方法；

(2) 当输入一个候选汉字之后, 余下的候选汉字被分为两组显示, 第一组是相应该输入汉字的相关词组的下一个汉字, 第二组是非联想的侯选汉字, 以便输入联想汉字或非联想汉字的任意侯选汉字的处理方法。

5 (3) 所述显示单元, 可用普通话或广东话的音标显示侯选汉字和部件的显示处理方法。

35、根据权利要求 34 的汉字输入方法, 其特征在于所述显示、处理方法还包括用于延续搜索下一组侯选汉字和部件的处理方法, 和利用所显示的侯选汉字及部件之后的点数表示待搜索侯选汉字及汉字部件组数的处理方法。

10 36、根据权利要求 34 的汉字输入方法, 其特征在于所述显示方法还包括可调整侯选汉字和部件的显示个数, 和侯选汉字从左至右排序显示或从中间到左、右两边排列显示的排列形式的显示处理方法。

37、根据权利要求 34 的汉字输入方法, 其特征在于所述键盘是以定位点为标志, 在键位上有序排列输入笔划的。

15 38、根据权利要求 34 的汉字输入方法, 其特征在于还包括所述虚拟键盘的键位上可对应显示侯选汉字和部件, 以便随虚拟键位输入的方法。

39、根据权利要求 34 的汉字输入方法, 其特征在于所述信息处理单元可以输出任意所需的汉字外部标准编码。

40. 根据权利要求 34 的汉字输入方法, 其特征在于在选择了一个侯选部件后, 可显示以所选部件为第一书写结构的复杂部件。

20 41、根据权利要求 34 的汉字输入方法, 其特征在于可用色标区分联想词组和非联想词组, 用另外色标区分简单部件和复杂部件。

42、根据权利要求 34 至 41 之一的汉字输入方法, 其特征在于该方法可应

用于普通计算机输入键盘和显示方法。

43、根据权利要求 34 至 41 之一的汉字输入方法，其特征在于该方法可应用于普通手机和普通电话拨号盘和相应的点阵显示方法。

44、根据权利要求 34 至 41 之一的汉字输入方法，其特征在于所述显示方法可应用于其它任何汉字输入方法中。

5

汉字字符输入的方法和系统

- 5 本发明属于汉字信息处理技术领域，是一种计算机汉字输入方法及其系统。

汉字犹如一个浩瀚的海洋，字量有几万个。许慎的《说文解字》收录有一万多个汉字。公元十九世纪中叶成书的《康熙字典》收录有五万多汉字。今天，绝大多数受过较高教育的人平均掌握几千个汉字。这几千个字构成了现代汉语
10 中的几万个词组和短语。

笔画是汉字的基本构形要素。汉字中大约包括三十多种笔画形式。汉字的笔画数目相差很大，可从一画到几十画。三十几种笔画也有大小、长短和位置的不同。

较早的机器输入汉字的方法是使用规定的编码。例如，汉字电报电码使用
15 4位数字为9999汉字进行编码。一位经过专门训练的人需要熟练记住所有汉字的编码及其相关的键位。这种方法的一个显著优点是可以盲打，其最大的缺点是操作员必须熟练记住大量的编码数字。每当操作员不知道某一汉字的编码时，必须查找有关内容。这样势必会大大减慢输入的速度。

以拉丁字母为书写系统的文字体系相对容易输入，因为其26个字母与打字
20 机键盘上的键位相一致。从手动的打字机键盘向计算机键盘的过度也只不过是一个简单的调整而已。但汉字的情况就完全不一样，把几万个汉字设定在西方26个字母键盘上就不那么简单。

汉字输入的的另一种方法是利用汉字字音作为一种中介，即使用汉字的语音符号输入汉字。普通话是现代汉语的标准音。不包括声调，普通话中有400
25 多个不同的音节。因此，普通话中有很多同音字。再加上汉语方言中发音的不同，依靠汉字的语音辨别许多汉字会有一定的困难。

显然，一种能够输入全部几万个汉字、能够为广大操作员所掌握的简单、方便、实用、可行的汉字输入方法将十分有益于汉字输入的需要。

美国专利第5,109,352号提出一套方案：

- 30 1)把汉字中的笔画归为少数几种基本笔画；

2) 按照人们书写汉字的笔画顺序序列输入汉字。按照第5, 109, 352号专利的设想, 输入汉字需要的笔画数很多。虽然汉字笔画的书写顺序有一定的规律, 但是人们书写具体汉字的笔画顺序有时可能有很大不同。对于含有笔画比较多的汉字, 人们书写顺序完全相同的情况并不多。352号专利运用容错的方式解决这一问题。但是过多地允许容错这一补救方法会削弱其功能的独特性。

此外, 当操作员不知道某一汉字的笔画顺序或笔画时, 尤其是当笔画的组合有不只一种的可能时, 操作员需要花费一定的时间去试着输入或查阅参考资料寻找。这样会大大影响输入速度。

使用部首的输入方法通常含有一种复杂的编码、严格的规则和固定的键盘分配。由于部首主要是为了对汉字进行分类, 并不能表述汉字的全貌, 有些汉字的主要结构特征则不能显示出来。因此, 这种系统通常比较难学、难用。

上述的汉字输入系统中编码不统一, 记忆量大, 难以将汉字的特性和规律同现代科学技术充分地结合起来。因此, 难学、难用, 难以推广和普及。显然, 社会需要一种简单、易学、易用的大众化的汉字输入方法。为了满足这一需要, 下面推出一种新研制开发的、理想的汉字输入解决方案。

本发明的主要目标是提供一种优化的汉字输入方法和系统, 它只需很短的学习和训练时间, 具有很高的输入效率, 而毋需任何外部参考。

本发明还有以下几个目标:

1. 通过保留操作员对输入词组的思维, 给目标装置提供完整的词组以作进一步的字处理, 从而保留中文的词组特征;

2. 通过采纳所输入的汉字或词组的频率, 而允许该系统自身针对不同的操作员或小组自动进行优化和修正;

3. 允许操作员使用部件替代笔画以进一步降低重码率, 更精细地分辨出不同的候选字;

4. 针对先前输入的笔画或部件, 系统通过计算部件在所有以该部件作为下一个书写结构的汉字累加频率, 从而得到对部件的排序;

5. 为中文输入系统提供许多与众不同的特色, 如毋需在词组之间输入空格, 毋需在键盘上贴标签, 以及特殊的虚拟键如“延伸键”及“智能键”。

在本发明中, 操作员和输入装置形成一个封闭的反馈回路。操作员对该系统所显示的信息作出反应, 并通过键盘或其它手段提供额外的信息或作出决

定, 操作员的反应在很大程度上取决于他(她)对汉字或由汉字组成的语言的认识程度。该输入装置同时也对操作员所输入的数据作出反应, 并显示更多的视觉信息。另外, 该装置在下一步操作中的表现会随着上一步操作的不同而变化。

该系统可以使用标准键盘或触摸屏作为主要的输入手段。其中大部分的操作同时适应于以上两种输入手段。键盘可以是单手或双手操作的, 但双手操作键盘更适合于快速输入, 而单手操作键盘则适合于另一只手需要放在原稿上的誊写工作。当用触摸屏作为输入手段时, 操作员可以用手指或触摸笔在适当的显示区触摸来替代键盘操作。

输入手段可以包括一个传统的键盘, 即典型的计算机或打字机键盘。传统的贴在键盘上的标签在这里并不重要。在本发明中可以不使用这些标签, 或把它们用作其它用途, 如可以用作拉丁字母、注音符号, 日文假名或韩文字符。

本发明所希望的最佳设计是利用键盘上的一个小突点使右手或左手定位。因为毋需键盘上的标签, 除了开始时使手放在适当位置外, 也就不需要操作员把目光移到键盘, 在这一所希望的最佳设计中, 八个手指放在键盘的主键行上, 而这主键行的每一个键则对应于八个笔画之一。大拇指可以自由地控制空格键, 而双手的小指则可以用来控制“SHIFT”键。这一特色与以往的汉字输入方法有明显的不同。这一手指位置的安排可以使操作员仅使用主键行就可以高速输入笔画。操作员也毋需额外的(昂贵及费时)打字课程来熟悉键盘的编排。

本发明也包括用于汉字输入的显示方式。如果需要, 显示区的其中一部分可以专门用来作为操作员与输入系统在输入汉字时的交互对话。如果当前的操作是输入汉字, 则显示区可以与屏幕上其它信息重叠, 如与本发明有关的显示信息可能与一个字处理软件工具重叠, 而该字处理软件需要利用本发明输入的汉字。本发明同时也包括在屏幕显示用抽象符号代表的笔画、部件以及所支持的字符集中某一汉字的手段。

该汉字输入系统包含由程序控制的数值处理器及其存贮手段。它们提供了高速完成本发明中的输入/反馈过程所需的设备。本发明也使用存放在存贮手段中的数据, 这些数据包括与单个汉字特征有关的信息、与单个汉语复合词(词组或短语)特征有关的信息、以及与汉字的部件有关的信息。

单个汉字的特征包括:

(1) 书写该汉字所用的笔画编码及顺序。笔画编码与所选择的笔画分

类方案相对应;

- (2) 该汉字在操作员所用的语言中出现在词组开头时的频率;
 - (3) 该汉字按书写顺序所用的部件;
 - (4) 该汉字在所属汉字子集的标记, 如简体字、繁体字、日文、韩文、
- 5 中小学生常用简体字、与简体字相应的繁体形式、俗语或一组随机的字符集。
- 与中文词组特征有关的信息包括:

- (1) 该词组在操作员所用的语言中所出现的频率;
- (2) 该词组在所有汉语词组的不同子集的成员标记, 如常用普通话词组、地名、人名以及俗语等。

- 10 汉字在该系统中的内部编码可以是任意的。该系统可以用传统的编码转化方法如码表查找法把内部编码换成外部所需的编码。在本发明所希望的最佳的设计中, 所支持的所有汉字都按一个“汉字表”中的序数来存放, 并可以通过码表查找法转换成各种不同的16比特外部标准编码。

以下将根据附图介绍本发明的原理

- 15 图1A-1C 是本发明中的笔画及笔画分类, 其中:

图1A 列出了用于书写汉字的简单笔画。它们被分成了29种基本形式, 再分类成8个及5个基本笔画。

图1B 列出了29个基本形式及简化后的8个基本笔画之间的对应关系。

- 20 图1C 列出了29个基本形式与进一步简化的5个基本笔画之间的对应关系。

图2 列举了汉字的构成, 箭头2.1指示合成后的汉字, 箭头2.2指示构成该汉字的简单部件, 箭头2.3指示相应的笔顺。

图3 是复杂部件的例子(箭头3.1所示)。它们由对应的简单部件构成(箭头3.2)。

- 25 图4 是本发明所适用的一个系统整体结构的框图。

图5 是图4中所用的一个显示设计的例子。图中的显示设计为使用双手操作键盘作为输入手段, 并为操作员提供反馈信息。

图6 是当用触摸屏作为输入手段的显示装置。图中是为操作员提供信息的各个区域。

- 30 图7 是本发明中相应程序的流程图。这一流程图显示了程序的粗略执

行步骤。这一程序对操作员选择汉字，输入笔画及使用特殊键作出反应。

图8 是本发明相应的程序所使用的基本数据结构。

图9 是组成一个汉字的分步显示。

图10 是本发明相应程序获取汉字词组中的候选字过程的流程图。

5 如图5所示，当该汉字输入系统启动时，它就把键盘上该系统所需部分作为虚拟键盘在屏幕上显示出来。虚拟键盘包括用于放置手指的主键行及它的上一行。这样的指法最适合于快速汉字输入。值得注意的是这种最佳指法使得不用在真正的键盘上再贴上标签。手指在键盘上的正确位置是把食指放在有一小突点的那个键上，这样不用贴标签也可以使手指定位。

10 按照本发明的设计，显示屏有每一个手指对应的笔画的标记。这种手指与笔画之间的对应关系是固定的，这样便于操作员掌握，输入笔画时也不用过多的思考，加快了输入速度。每个笔画的影像显示在虚拟键盘所对应那个键的角上。如果是触摸屏，每个笔画也按同样方法显示，而输入则可以通过触摸该笔画的影像来完成。触摸屏的另外一种输入方法是在专门为此设计的一个显示区
15 内画出这一笔画。

操作员一个笔画接一个笔画地输入，每当输入一个笔画，汉字输入系统就在显示屏的输入区内把笔画累积起来。该系统所指的笔画实际上是真正汉字笔画的抽象分类。按照本发明所希望的最佳设计，汉字中的29个最基本的笔画按图1B所示被划分成8个抽象笔画。但也可以选择更多或更少的基本及抽象笔画
20 作为替代方案。

当输入每个笔画时，所显示的候选汉字会随之更新，更新后的候选字就是与所输入的笔画顺序相对应的出现频率最高的那些字。当出现所需的汉字时，操作员只需选取这个汉字(即按下那个汉字在虚拟键盘上所对应的键)就达到了目的，操作员可以通过学习手指位置与笔画之间的对应关系以及所需汉字
25 所对应的键与附近笔画键之间的相对位置来选择该汉字，使眼睛不需要看着键盘。

显示区里的候选字是可以直接通过键盘来选取的，因为显示屏上候选汉字是直接显示在虚拟键盘的键上。本发明所希望的最佳设计是把这些汉字显示在主键行的上一行。操作员可以通过按下该汉字所对应的键来选取这一汉字。操
30 作员的眼睛就可以(而最好也应该是)继续停留在显示屏上。操作员通过这些汉

字所对应的键与主键行各键之间的对应关系来定位。显示区内可以显示多达10个候选字。在所希望的最佳设计中，操作员可以调整候选字的显示个数及排列方法，显示个数可以是在5到10之间，排列方法可以是从左到右，从右到左或从中间向两边排列。虽然把尽可能多的候选字从左到右排列对西方人是显而易见的，但从中间到两边这一排列却正反映了我们的一大发现，即中文操作员因为可以从一小堆汉字中迅速找出所需的汉字，因而他们更喜欢把较少的候选汉字按从中间到两边的排列。

如图5所示，在所希望的最佳设计中在主键上提供了一个“延伸键”47。按下这一“延伸键”，显示区内会出现下一组频率最高的候选字。这一“延伸键”在虚拟键盘上的影像也包括了一组小点49，这组小点代表了还剩多少组以所输入的笔画开头的候选字。这样就允许操作员选择输入更多的笔画来减少候选字的个数，或选择用“延伸键”来浏览各组候选汉字，每一小点代表一组候选汉字。

本发明所希望的最佳设计也包括一个表示任意笔画的“智能键”46。它具有模糊输入的功能。这一特色在当操作员不知道或不能肯定某一汉字在这一系统中的笔画分类或正确笔顺时特别有用。

本发明所希望的最佳设计中另一个特色是把“最终必须显示”的候选字排在通常的按频率排列的其它候选字前面。在大多数情况下，简单的部件，即笔画数较少的那些通常也是最常用的，但是也有不少例外。有些笔画较少的汉字比以相同笔画开始的更为复杂的汉字远为罕见。如果所支持的字符集中包含了那些很少作为单独汉字来用的部件时，这种情况更容易发生。这时操作员不能再输入更多的笔画，因为所需的汉字再没更多的笔画。如果输入更多的笔画只会把该汉字从候选字中除去。在这种情况下，本发明的处理方法是把所有与所输入笔画完全相匹配的那些汉字排在其它汉字前面。这样操作员就可以看到这些汉字，这些汉字可能会被安排在第一组候选汉字中，也可能被放在某一特定候选字组的开头(即可以用“延伸键”来浏览的那些汉字)。必须注意到，这时操作员已经知道他(她)已输入了那个汉字的所有笔画，所以也就知道只能用“延伸键”来找所需的汉字。

本发明巧妙地利用了部件作为一组笔画的替代，并作为一种手段来更精细地筛选候选汉字。这一特色对那些复杂字和罕见字尤为有效。这一组部件是通

过对所支持的特定的汉字字符集中所有汉字的部件结构及书写顺序进行分析而得来的。

图2显示了本发明所希望的最佳设计中选择汉字时的图像及笔画顺序。

在以后的讨论中，我们把单纯由最基本的笔画组成的部件称为“简单部件”，而把那些由其它部件所组成的部件称为“复杂部件”，图3是几个图2中所示的复杂部件的例子。

按照本发明，显示区所显示的下一个所要书写部件的次序是按照与该部件对应的所有汉字的累加频率来排定的。当给定先前已经选择的笔画或部件时，系统通过对候选汉字组中所有汉字的每一个部件进行频率累加而得到排列次序，通过计算与当前所输入的笔画或部件相匹配的所有汉字中每一个部件的累积频率，系统就可以给出可以最快找到所需汉字的那些部件。这与传统汉字输入方法中把偏旁部首固定排列的设计相比，是一个巨大的进步。

在本发明所希望的最佳设计中，部件被显示在虚拟键盘的主键行上。部件的选取可以通过先按下“SHIFT”键，再按下该部件所对应的那个键来完成。按下“延伸键”则可以显示额外的部件。与显示候选汉字组数目的情况类似，剩下的没有显示的部件个数以小点的形式显示在“延伸键”上。当然也可以用其它方式来显示，如可以将候选字显示在主键行上，当按下“SHIFT”键时，主键行就用来显示部件。

当操作员选择了一个部件，如果在输入区中含有先前输入的笔画，所选择的部件则替代了前面的笔画。现在就只剩下那些以该部件作为第一个书写结构的候选字。选择部件可以使得候选字数目大为减少。这是由于虽然许多汉字的前几个笔画相同，但部件的个数却比笔画的组合数要多，一到二个笔画再加上一个部件就可以有效地把不同的汉字区分开来。例如，输入三个笔画后可能还存在几百个候选字及几十个部件，但如果选择了一个部件，则候选字的范围收窄到以该部件作为第一个书写结构的那些汉字。这些候选字的个数大大小于再输入一个笔画所得到的候选字数。另外，还有许多汉字的第一个书写结构含有许多笔画，如果单纯地使用笔画，就需要把该部件的所有笔画都按顺序输入才能把这些候选汉字区分开来，因而采用部件可以戏剧性地减少所需输入的笔画数，尤其是那些较为罕见的汉字。

本发明的另一个特点与部件的复杂结构有关。复杂结构是指有些部件包含

了另外一些部件。在所希望的最佳设计中，一旦操作员选择了一个部件，该系统就找出所有以所选部件作为第一个书写结构的复杂部件，并把它们排在可能是候选字下一个书写结构的那些部件前面，且用颜色或其它手段把两者分开。

这样的大“包含”部件的排列次序是基于把所有以这些“包含”部件作为第一

5 个书写结构的汉字频率进行累加而得到的。

当选择了一个部件后，操作员通常有以下几种选择：

(1) 如果所需的汉字已经出现，就可以直接选择它；

(2) 选择了一个大“包含”部件作为第一个书写结构；

(3) 如果先前所选择的部件是所需汉字的第一个书写结构，则可以选择
10 该汉字的下一个结构；

(4) 输入下一个笔画。

除此之外，还可以用“延伸键”去浏览额外的同类候选字，候选字的个数由“延伸键”上的小点来代表。值得注意的是，所有中国人毋需任何指导就可以使用该系统。操作员对偏旁部首的结构和分类知识毋需如学者般丰富，也毋
15 需知道偏旁部首在键盘上的复杂对应关系，操作员只需知道怎样书写该汉字以及怎样把笔画重复地组合成部件。

由于该汉字输入系统反映了汉字的自然书写习惯(部件的书写顺序，部件中可能包含更简单的部件，每个部件由笔画组成)，所以整个系统与真正的汉字书写习惯相吻合。

20 当输入部件后再输入笔画时，候选字、大“包含”部件及下一个书写部件都受到所输入笔画的制约，每输入一个笔画通常会减少上述三类中每一类的候选数。在实际操作中，操作员已在心目中想好了所要输入的汉字，并将目光停留在显示屏上，同时提供额外的笔画和部件直到所需的汉字出现为止。

一旦操作员输入了一个汉字，候选字就被分成两组显示，第一组是词组的
25 下一个汉字，这些词组是一个任意大小的词组字典中的一个子集。在本发明所希望的设计中，这个字典含有大约80,000个词组、短语、成语和人名。当操作员输入下一个汉字的笔画或部件时，第一组中的候选字数就会迅速减少，第二组则与先前没有输入汉字时的候选字完全相同。

在本发明的一个设计中，词组联想的候选汉字以红色显示(图5，箭头54)，
30 而不是词组联想的那些候选字则用黑色显示(图5，箭头53)。换句话说，词组

联想的候选字排在不是词组联想的候选字前面，用不同的颜色把两种候选字区分开来。在许多情况下，词组联想的候选字会多于一个显示页，前面所述的“延伸键”就被用来显示下一组候选汉字。“延伸键”上的小点(图5，箭头49)也用红色和黑色显示，这样操作员就可以立即知道还剩下多少组词组联想的或不是联想的候选字。当输入笔画或部件时，这两种候选字的数目都会减少。由于这时剩下的词组仅仅是汉字总数的一小部分，因而随着笔画或部件的继续输入，候选字列的数目会迅速递减。

如果这时选择一个与词组相关的候选字(图5中的红色显示，箭头54)，就把该汉字作为当前词组的下一个汉字输入，而如果选择不是词组联想的候选字(图5中黑色显示，箭头53)，输入系统就开始了一个临时性的新词组。之所以是临时性的是因为在所希望的最佳设计中，操作员可以为词组字典定义一个新词组。尽管词组字典可能很大，却不能包含所有词组如人名或技术术语。如果这时操作员没有示意要加入新词组，词组联想的候选字就是词组字典中那些含有先前输入的汉字的词组。

本发明的另一个特点是毋需在词组之间输入空格，操作员可以输入一个空格来表示当前词组的结尾，并把该词组送入应用程序(如果操作员在输入系统设定，也可以把词组分隔符连同词组一并送入应用程序)。但是操作员也可以不输入分隔符，本系统把词组放在缓冲区中，并把当前的词组用红色显示(也可以选用另一种与缓冲区其它词组不同的颜色)。当操作员选择了一个与词组不相关的汉字(黑色显示)时，上一个词组就自动结束了。当缓冲区被存满或操作员输入一个空格来明确表明结束当前词组并移走其中的汉字时，缓冲区中的汉字就按应有的次序被送到指定的应用程序。

本发明的另一个特色是它不仅可以用笔画和部件显示候选字，而且还能用普通话及广东话的发音来显示候选字。这一特色允许操作员输入罗马字母或注音符号来代替笔画，键盘上的罗马字母用于在中国流行的拼音输入，而注音符号则用于在台湾流行的注音输入。注音符号在台湾通常印在键盘上。操作员可以结合“CTRL-SHIFT”，“ALT-SHIFT”及键盘上的罗马字母和注音符号来输入这些“字符”。本发明也适用于使用其它键作同样的用途。罗马字母或注音符号就这样被输入到虚拟键盘的输入区，并被用于把候选字的范围缩小到其标准发音也与所输入的音标相匹配。

当操作员不能确定某一汉字的笔画分类及顺序时,这一特色提供了选取所需汉字的另一个途径。因为部件的显示次序是按照候选字的频率来排列的,在许多情况下仅输入一个音标字符就可以找出所需的部件而毋需再输入任何笔画。在很多情况下,输入第一个音标后再加上该字的起始笔画(这一起始笔画是很少会混淆的)后,即使不能找到所需的汉字,也能保证该系统能显示一个有用的部件。

综上所述,本发明提出了一种计算机汉字输入系统,其中包括:

- a. 依汉字笔序利用笔划码元输入汉字的键盘单元或虚拟键盘单元;
- b. 汉字信息处理单元; 及
- 10 c. 汉字及其它相关信息的显示单元; 其主要特征在于还包括:
 - (1) 在笔划码元输入操作中代替任意笔划码元的智能笔划键处理单元;
 - (2) 随着笔划码元的输入,在显示单元上逐次更新显示预测引导性的依使用频度排序的相应候选汉字及相应候选待用汉字部件的处理单元;
 - (3) 随实际输入汉字、词的频度而随机改变相应于输入笔划码元的汉字、
 - 15 词频度排序及相应汉字部件频度排序的随机频度排序显示的处理单元;
 - (4) 当输入一个候选汉字之后,余下的候选字被分为两组显示,第一组是相应该输入汉字的相关词组的下一个汉字,第二组是与在先候选字完全相同的汉字,并用色标区分这两组,以便不用空格即可输入联想汉字及非联想汉字的任意候选汉字的处理单元。

- 20 所述显示、处理单元还包括用于延续搜索的延伸键处理单元及显示排序汉字及部件后的点数表示待搜索候选汉字及汉字部件组数的处理单元。

所述显示单元还包括从左至右排序显示或从中间到左、右两边的显示处理单元。

- 25 所述汉字输入系统,可应用于普通计算机的输入键盘和显示系统中,还可应用于普通电话拨号盘和相应的点阵显示单元,还可应用于普通手机拨号盘和相应的点阵显示单元。

所述显示处理单元可应用于其它任何汉字输入系统中。

- 30 本发明还提出了一种计算机汉字输入方法,该方法利用普通键盘单元或虚拟键盘单元,并利用汉字信息处理单元及显示单元,依汉字笔序利用笔划码元输入汉字,其特征主要在于:

(1) 在笔划码元输入操作中代替任意笔划码元的智能笔划键处理方法;
(2) 随着笔划码元的输入, 在显示单元上逐次更新显示预测引导性的依使用

频度排序的相应候选汉字及相应候选待用汉字部件的处理方法;

- 5 (3) 随实际输入汉字、词的频度而随机改变相应于输入笔划码元的汉字、频度排序及相应汉字部件频度排序的随机频度排序显示的处理方法;
(4) 当输入一个候选汉字之后, 余下的候选字被分为两组显示, 第一组是相应

该输入汉字的相关词组的下一个汉字, 第二组是与在先候选字完全相同的
10 汉字, 并用色标区分这两组, 以便不用空格即可输入联想汉字及非联想汉字的任意候选汉字的处理方法。

所述显示、处理方法还包括用于延续搜索的延伸键处理方法及显示排序汉字及部件后的点数表示待搜索候选汉字及汉字部件组数的处理方法。

所述显示方法还包括从左至右排序显示或从中间到左、右两边的显示处理
15 方法。

所述汉字输入方法, 可应用于普通计算机的输入键盘和显示系统中, 还可应用于普通电话拨号盘和相应的点阵显示单元, 还可应用于普通手机拨号盘和相应的点阵显示单元。

所述显示处理方法可应用于其它任何汉字输入系统中。

20 下面结合附图对本发明作更加详细的说明

图4是使用本发明装置的整体结构框图。该基本装置由输入手段10(可能是一个键盘及触摸屏), 处理器20(包括传统的中央处理器, 它执行在存贮手段30中存放的程序指令, 中央处理器也读取存储器中本发明所用的数据并存放中间结果), 以及传统的显示手段40(包括全程可控地址的显示器如CRT, 液晶显示器或其它类似的显示装置), 也可以附加一个发声装置50如扬声器, 当输入出现错误时, 它可以发声引起操作员的注意, 由本发明可以不难派生出各种不同的设计方案。这些不同的设计方案中的不同部分可以轻易地在本发明的范围内相互互换。中央处理器20及存贮器30提供了本发明所需的执行所存贮程序的手段, 该程序完成详细的的处理过程。存贮器30同样也存放本发明所需的中间结果。
25
30 果。在所希望的最佳设计中, 所有的信息通过一个或多个重叠的窗口(如图5和

图6)显示给操作员。而这些窗口被放置在任何应用程序(如字处理软件)的上面。

显示区域包括一个虚拟键盘区(图5, 41)。虚拟键盘的排列与真实键盘相吻合。输入元素缓冲区52用来显示操作员所输入的笔画及部件。字符缓冲区44
5 则用来显示操作员已选择的汉字。这些不同的显示区也可以用其它的排列方法, 如把这些显示区用在不同的窗口来显示, 也可以把汉字、笔画及部件显示在应用程序窗口中。

虚拟键盘包括指示物48。它对应于传统键盘主键行上的小突点, 用于引导手指放在正确位置, 这样就不需要把目光移向键盘。虚拟键盘也可以显示与真
10 实键盘对应的罗马字母。

键盘主键行包括基本笔画51的抽象表达, 每一个笔画对应于一个键。在所希望的最佳设计中, 中间的四个键对应于非常熟悉及常用的笔画“横”, “竖”, “撇”, 及“点”。剩余的键则对应于其它四个笔画“横折弯钩”, “竖折”, “横折”, 及“弯钩”。笔画和手指位置是一一对应的, 这样经过反复使用后,
15 操作员可以迅速地输入笔画, 包括单一的笔画或更为常见的两到三个笔画顺序。本发明中的部件也显示在主键行上。所显示的那些部件是先前输入的笔画及部件的函数, 它由下面所述的步骤确定。在本发明的一个具体设计例中, 简单部件55以蓝色显示, 而复杂部件56则以绿色显示, 颜色的选择并不重要, 只要能区分它们就可以了。

20 操作员通过按下一个笔画所对应的键来输入笔画。当用触摸屏作为输入手段时, 笔画的输入则通过触摸显示区中代表该笔画的部分, 或通过专门为此设计的区域中画出这一笔画。

部件的输入是通过使用“SHIFT”键, 同时按下与该部件相对应的键来完成。当用触摸屏作为输入手段时, 部件的输入则是通过触摸该部件所在的显示
25 区来完成。在本发明所在的最佳设计中, 主键行上同时也包含一个智能键(图5, 46)。在智能键上标有一问号。当操作员按下此键时, 它代表8个笔画中的任何一个。

虚拟键盘主键行的上一行(图5, 42)显示与先前输入的笔画或部件相匹配的候选字。这些候选字按出现频率的高低来排列。可以通过按下某一汉字所对
30 应的键来选择该汉字。操作员通过显示屏上与主键行的对应关系来完成操作,

例如, 如果所需的汉字出现在第一个键上, 操作员可以立即知道该键是主键行左手小指所在键的上方。这样的视觉关系使得操作员没有必要把目光从屏幕上移开。当使用触摸屏输入时, 即可以简单地通过触摸该汉字所在的显示区域。与先前输入的汉字相关的那些词组候选字以红色显示⁵⁴, 而非相关汉字则以黑色显示⁵³。词组候选字排在非词组候选字的前面。

在所希望的最佳设计中, 主键行上的一个键被设计成延伸键(图5, 47), 按下该键使得系统显示下一组候选字。显示屏上该键下方还有一组小点(49), 用来代表还剩下多少个与所输入的汉字、笔画及部件相匹配的候选字及部件。这些小点用颜色来表示多少候选汉字是词组联想(红色⁵⁴)或不是词组联想(黑色⁵³)。同样地也用颜色表示多少是复杂部件(绿色⁵⁶)及简单部件(蓝色⁵⁵)。

关于本发明中相应程序的数据结构, 我们把各种数据元素解释成不同的数值存放位置。这些数据元素通常以二进制形式存放, 以便特定的中央处理器能方便地存取及处理。

汉字表(8. 1)是一个多重入口的数据结构, 每一个汉字对应一个入口(参考图8)。本发明的某一特定设计所支持的汉字数目取决于其所处的应用环境, 通常从几千到多于二万不等。在某些情况下, 为了支持某些汉字使用多种笔画顺序, 这些汉字可能在汉字表中有多个入口。

这一汉字表中的汉字是基于这些汉字在文章中出现的频率来排列的。在本发明所希望的最佳设计中, 汉字的真实使用频率可以通过计算操作员对该汉字的实用使用次数来计算并反映在汉字表中, 在这种情况下, 汉字表中汉字的次序仍然按频率来排列。应该注意这里的频率并不是相对频率或次序, 而是绝对频率, 即在N个词组中(N是变量)该汉字做为词组的第一个字所出现的次数。

每个汉字表的入口包含以下数据:

1. 该汉字的笔画顺序(图8, 8. 2)。在所希望的最佳设计中, 每个笔画被赋予一个1到8之间的数值。0则被用来代表笔画顺序的结束。一个汉字所需存储的笔画数一般大大小于36这个所有汉字中最大笔画数, 由于每个汉字只需存放在足以把该汉字从某一组中区分开来所需的笔画, 所以需要选择一个所需存放的最大笔画数。在本发明所希望的具体体现中, 这个最大的笔画数是24。它主要用于节省计算机的内存空间。

2. 对每个汉字的前两个部件进行编码(图8, 8. 2), 每个部件被赋予一个

任意值。在本发明所希望的具体体现中，有多于1000个部件，但是我们发现只需把汉字的第一个最大的部件及第二部件个进行编码就足以把汉字划分成很小的组别。

3. 汉字表包含的汉字的频率(图8, 8.4)以及其它一些可选的但非常有用的汉字特征信息(8.5)。这些信息可以包括该汉字在不同编码方案中的编码及在不同字符集中的成员属性，也可以包含该汉字所代表的音标编码(8.6)。这些音标可以做为笔画和部件以外的手段来输入汉字。

4. 汉字表包括一个对应于词组表(8.11)入口的指针。词组表存放着可以与该汉字组成词组的那些字。词组表将在下面介绍。

10 第二个重要的数据结构是部件表(图8, 8.8)。部件表中有所有部件的入口，每个入口包括该部件的笔画顺序(8.9)，但也可能包括该部件的第一个书写部件(8.10)。部件表的每一个入口只允许一个部件，但是该部件可能由另一个部件组成，同时该部件也可能被用来组成其它部件。

第三个主要的数据结构是词组表(8.11)。在所希望的具体体现中，词组表
15 允许存贮几十万个复合词组(至少两个汉字)及短语，目前存放着常用的80,000个词组，它们构成汉语词组的绝大部分。词组表本身是这些词组的一个链表，为了节省存贮空间，词组中的第一个汉字没有被存放，因为词组表是通过汉字表中以该汉字开头的所有词组的指针来读取的，所以词组表中的第一汉字已经包含在汉字表中。在词组的汉字后面是一个非字符编码，该非字符编码可以是
20 词组的频率。词组的结束也可以用下一个字组的开始来蕴含，而词组的开始则由汉字表中的另一个入口来定义，这样可以进一步压缩存贮空间。

图7是一个抽象的流程图，它代表了系统对操作员输入的响应。除了前面所述的主要数据结构以外，系统还使用用于存放先前输入的汉字的“字符缓冲区”以及用于存放先前输入的笔画及部件的“输入元素缓冲区”，这两个缓冲区在程序开始运行时就被初始化了。
25

从所存贮程序运行的观点来看，操作员的输入，无论是键盘、触摸屏、或是其它类似的手段，在概念上是完全等同的。在每一种情况下，操作员提供候选字的选择，部件的选择，笔画的选择(包括“智能键”做为虚拟键)，“延伸键”的请求，及输入“空格键”或“退格键”。图7的流程图显示了每一种情况下的执行步骤。
30

如果是笔画的选择,所选择的笔画就被放在输入元素缓冲区的末端,然后获取候选字并更新显示屏。获取候选字的详细步骤将在下面讨论。在通常情况下,候选字的数目可能达到预先设定的有关相关汉字、非相关汉字、复杂部件及简单部件的最大数目。如果在字符缓冲区中没有先前输入的汉字,则词组联想的候选字就不会出现,同样,如果在输入元素缓冲区中没有先前输入的部件,候选的复杂部件也不会出现。

按下“延伸键”会导致程序显示下一组候选字,而选择一个部件会导致程序首先确定操作员是否选择了一个复杂部件。如果是,当前所选择的部件就代替了先前输入的部件并将随后的所有笔画从输入元素缓冲区中删除。如果所选择的只是一个简单的部件,则会将所有的笔画(不包括部件)从输入元素缓冲区中删除并将所选的部件加在缓冲区的末端。程序将重新获取新的候选字并重新更新显示区。值得注意的是,这样的处理方法使得输入元素缓冲区只保留汉字的第一个部件(如果有的话),然后才是笔画,因为输入部件时就把后面的笔画删除了,因而部件绝不可能放在笔画之后。

选择一个候选字会导致程序首先确定所选择的是不是一个与词组联想的汉字(在设计例中的红色显示)。如果是,该汉字则被添加在字符缓冲区的末端,而如果不是一个词组联想的汉字,则首先在字符缓冲区中加入一个词组分隔符,再加入所选择的汉字。这两种情况都使得输入元素缓冲区中的内容被清除,显示器上的内容也被更新。

值得注意的是,这样处理的结果是字符缓冲区中可能包括由分隔符隔开的词组串。这是本发明的一个特色,这样就可以把句子中的各个词组区分开来,从而允许在词组的层次上,而不是在互不相关的的汉字层次上作字处理。当汉字和词组在字符缓冲区中连续地组合后,程序就可以把缓冲区中的所有汉字作为整个词组送到应用程序中。根据不同的需要,是否把词组分隔符连同词组一起送到应用程序这一参数可以在系统中设置,这要取决于应用程序是否能够处理这一分隔符。这样就避免了需要在不同词组之间输入空格符或其它类似机制来明显地区分它们。

“空格键”用来明显地发出结束当前词组的信号。在大多数情况下,单个汉字本身可能是词组,或是词组的第一个字。“空格键”用来明确地通知应用程序该汉字确实是一个词组的结束。如果有许多词组或短语以该汉字开头时,

这一特色尤其重要。由于通常笔画及部件的输入如果不是把全部,也会把大部分与词组相关的的汉字从候选字中排除并找出所需的非相关汉字,因此使用“空格键”通常只是操作员的一种偏爱,而使用颜色或其它方法来区分是否是词组的继续(红色)或是新词组的开始(黑色)则更增强了区分的效果。

- 5 “退格键”实际上是通过从相应的缓冲区中删去笔画、部件或汉字来取消先前的输入。当各缓冲区的内容被更新后,其候选字也被更新并在显示区中显示。如果先前所输入的是一个用于显示下一组候选字的“延伸键”,则输入“退格键”后就毋需再重新计算候选字而只是把上一组候选字重新显示。

前面已经对获取候选字的过程进行介绍。这一过程是本发明的核心部分,
10 现在利用图10再进行介绍。如前所述,存贮器中存放的程序和数据的起始过程会使得该系统处于某一状态,这一状态包含以下数据:

1. 当前词组中的汉字,如果有的话;
2. 输入元素缓冲区,它包含先前输入的笔画、部件、及可选的罗马和注音符号;
- 15 3. 当前所显示的那一组候选字,使用“延伸键”可以得到下一组候选字。而获取候选字的过程总是在这一状态下进行。

这一过程的第一步(图10, 10.1)是清除临时的数据结构,这一数据结构用于存放部件(图8, 8.13)及复合部件(图8, 8.14)的频率表,它有N个入口,这里N是部件编码的个数,所有的入口在初始时都置零。

- 20 第二步(图10, 10.2)是把输入元素缓冲区中的部件展开成笔画。其作法是检查缓冲区中的每一个元素,如果该元素是一个部件,则把部件表中该部件的笔画顺序复制到“笔画匹配”缓冲区中(图8, 8.15),如果在输入元素缓冲区中遇到笔画,则把笔画也复制到“笔画匹配”缓冲区中,如果遇到音标字符,该字符就被复制到“音标字符”缓冲区中(图8, 8.16)。在这一过程的最后,
25 “笔画匹配”缓冲区中存放的是所需汉字的笔画顺序,而“音标字符”缓冲区中可能含有对应于该汉字音标的那些字符。输入元素缓冲区中的内容并没有改变,如果操作员先前已经选择了部件,缓冲区中第一个及第二个元素可能是部件的编码。

- 这一过程的第三步(图10, 10.3)是检查汉字表(图8, 8.1)中的汉字是否与
30 所输入的数据相匹配。注意汉字表是以频率排序的,表中的第一个入口就是最

常用的汉字。如果以下的条件满足，则称该汉字是匹配的，即是候选字：

1. “笔画匹配”缓冲区中的笔画必须与所对应的汉字相匹配，但在比较时，“智能键”可以与任何笔画相匹配；

2. 如果输入元素缓冲区的第一个元素是部件编码，则候选字的第一个部件必须与该部件相匹配；

3. 如果输入元素缓冲区中第二个元素也是一个部件，候选汉字的第二个部件也必须与该部件相匹配；

4. 如果“音标字符”缓冲区中含有该汉字的音标字符，候选字的音标也必须与这些音标字符匹配。

10 由于这一目标，当两个部件的编码完全相同(即相等)时，就认为是匹配的。如果汉字的部件是复合部件，当该复合部件的第一个书写结构的编码与另一个部件的编码相同时，它们也被认为是匹配的。部件包含有一个可选的部件以及额外的笔画，如果部件中包含有另一个部件，则被包含的部件也可能再包含一个部件，这样递归到不再包含部件为止，即到只包含纯笔画的简单部件。匹配的过程是通过不断地递归并比较部件是否匹配，直到部件只含笔画为止，在这种情况下，部件不是一个匹配。

如果汉字并不匹配，这一过程就转入下一个入口再与另一个汉字相比较，如果该汉字是笔画相同的完全匹配，这个汉字就被复制到“完全匹配”列中。如果不是完全匹配，即汉字的笔画数多于需要匹配的笔画数，即该汉字就被复制20 到候选字列中。之所以单独计算机完全匹配汉字是因为这些汉字将会排在所显示的候选字的最前面。

另外，如果是不完全匹配，该汉字的频率就被按照“下一个”部件累加起来。前面讲过每个汉字的记录中有第一及第二个部件，如果输入元素缓冲区中正好有一个部件，“下一个”则是该汉字的第二个部件，而如果缓冲区中没有任何部件，“下一个”部件则指的是该汉字的第一个部件。部件的频率是通过25 把它累加到先前得到的部件频率上来获取的。另外，对包含在该部件中的每一个部件把汉字的频率累加起来。通过对每一个所匹配的汉字的频率累加后，所得到的累加结果反映了部件在相匹配的汉字中出现的频率。而如果把把这些部件按照累加频率排序后，操作员所得到的部件已经是按照找出所需汉字的概率大小30 小排列好了。注意在所有相匹配的汉字中，那些累加频率为零的部件不会出现

在后面,也不会把操作员引向所需的汉字,因而没有必要显示出来。如果字符集中所有汉字都是匹配的(因为还没有输入任何笔画),所显示的部件仍然是按照找出所需汉字的概率大小排列的。如果汉字表中的频率被调整到操作员对该汉字的实际使用频率,部件的排列次序就会自动反映出这一变化。

- 5 除了对“下一个”部件进行频率累加外,当输入元素缓冲区中含有单个部件时,第一个复合部件的频率也被累加起来,这一累加只有在部件与输入元素缓冲区中的第一个部件不是完全匹配时才进行。因为该汉字本身是一个匹配,它的部件及任何与输入元素缓冲区中部件不匹配的其它部件代表了复合部件,通过累加复合部件的频率,就可以把它们提供给操作员供他选择。这些累加频率组成“复合部件频率”表。

当整个汉字表被搜索完毕,将会有:

- (1) 包含了所有完全匹配汉字的“完全匹配”列;
- (2) 以频率排序的其它匹配;
- (3) “下一个”部件频率表,表中是相关部件的非需频率值;
- 15 (4) “复合部件”频率表,表中是所有相关部件的非零频率值。

这些频率表被用作进一步处理,然后得到一个部件编码—部件频率的序列,并按照频率进行排序。最后用这些部件编码组成按频率排序的部件及复合部件表。

- 20 这一过程的第四步(图10, 10.4)是产生词组联想的候选字,只有当字符缓冲区中含有先前输入的当前词组中的汉字时,这一步才进行计算。这一过程从当前词组的第一个汉字开始,首先在汉字表中找到这一汉字在词组表中的词组入口,并对所有有关词组进行分析。对每一词组首先检查随后的汉字是否与当前词组中随后的汉字相匹配,如果不匹配就移到下一个词组。

- 如果是匹配的,就利用词组字典中的下一个汉字获得该汉字的笔画顺序。
- 25 然后再检查该汉字的笔画、部件或音标符号是否匹配,匹配的方法与前面所述的完全相同。如果词组是匹配的,就把这一汉字加到词组联想的候选字列中。但在加入词组之前,还要检查一下该汉字是否已经在候选字列中,如果是就不再加入。当这一过程对所有的词组搜索完毕后,所产生的汉字列就是词组联想的候选字列。

- 30 显示过程(图10, 10.5)就会更新在显示屏上的候选汉字及部件。通常候选

汉字和部件往往会多于所能显示的数目，至于哪些会出现在显示屏上则取决于“延伸键”被按下多少次。系统中有一计数器记录哪一组候选字或部件会出现在显示屏上。

5 显示的排列次序首先是词组联想的候选字，然后是完全匹配的候选字，最后是其它候选字。部件的显示次序首先是复合部件的绿色显示，然后是其余部件的蓝色显示。

虽然上述已就本发明的技术特征及实施例作了详细的介绍，但是其现有的功能和特征在本技术领域的实际应用过程中还可以作进一步的修改或调整。所有这些修改与调整都应属于本发明的精神和所附权利要求限定的范围之内。

字元	国标	部 件	笔画数	笔 划
啊	B0A1	口 ㄩ 可	10	ㄣ ㄣ ㄣ ㄣ ㄣ ㄣ
阿	B0A2	ㄩ 可	7	ㄣ ㄣ ㄣ ㄣ
埃	B0A3	ㄣ 厶 矢	10	ㄣ ㄣ 厶 ㄣ \ / ㄣ ㄣ \ /
挨	B0A4	扌 厶 矢	10	ㄣ ㄣ 厶 ㄣ \ / ㄣ ㄣ \ /
哎	B0A5	口 ㄣ ㄣ	8	ㄣ ㄣ ㄣ / \
唉	B0A6	口 厶 矢	10	ㄣ ㄣ 厶 ㄣ \ / ㄣ ㄣ \ /
哀	B0A7	ㄣ 口 ㄣ	9	\ / ㄣ ㄣ ㄣ \ / ㄣ \ /
皑	B0A8	白 山 己	11	/ ㄣ ㄣ ㄣ ㄣ ㄣ ㄣ ㄣ

↑
2.1 汉字

↑
2.2 部件组成

↑
2.3 笔画组成

图 2

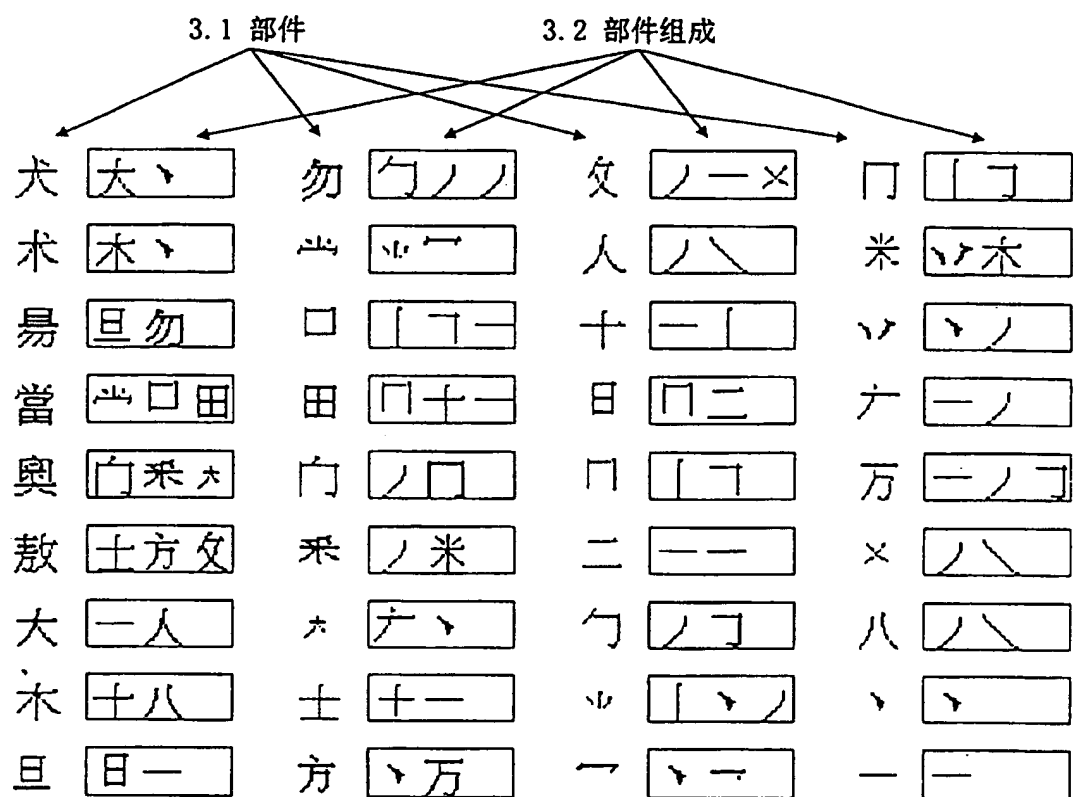


图 3

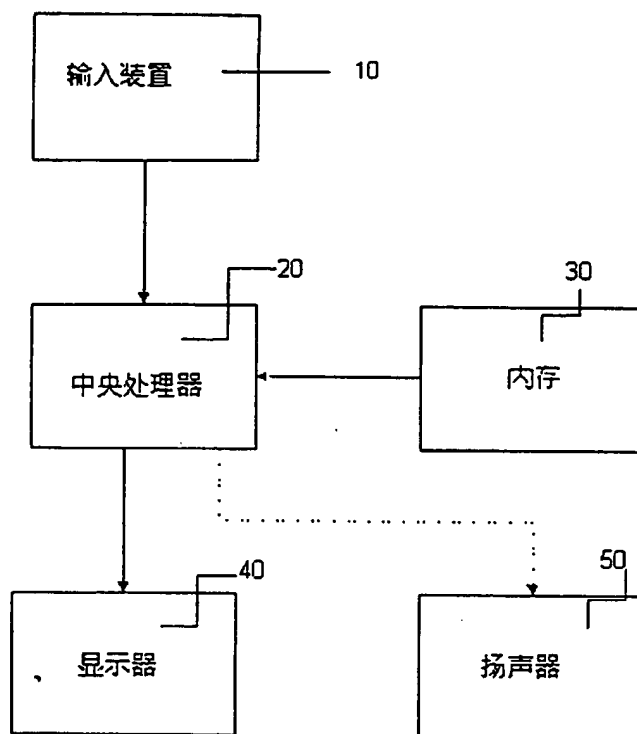


图 4

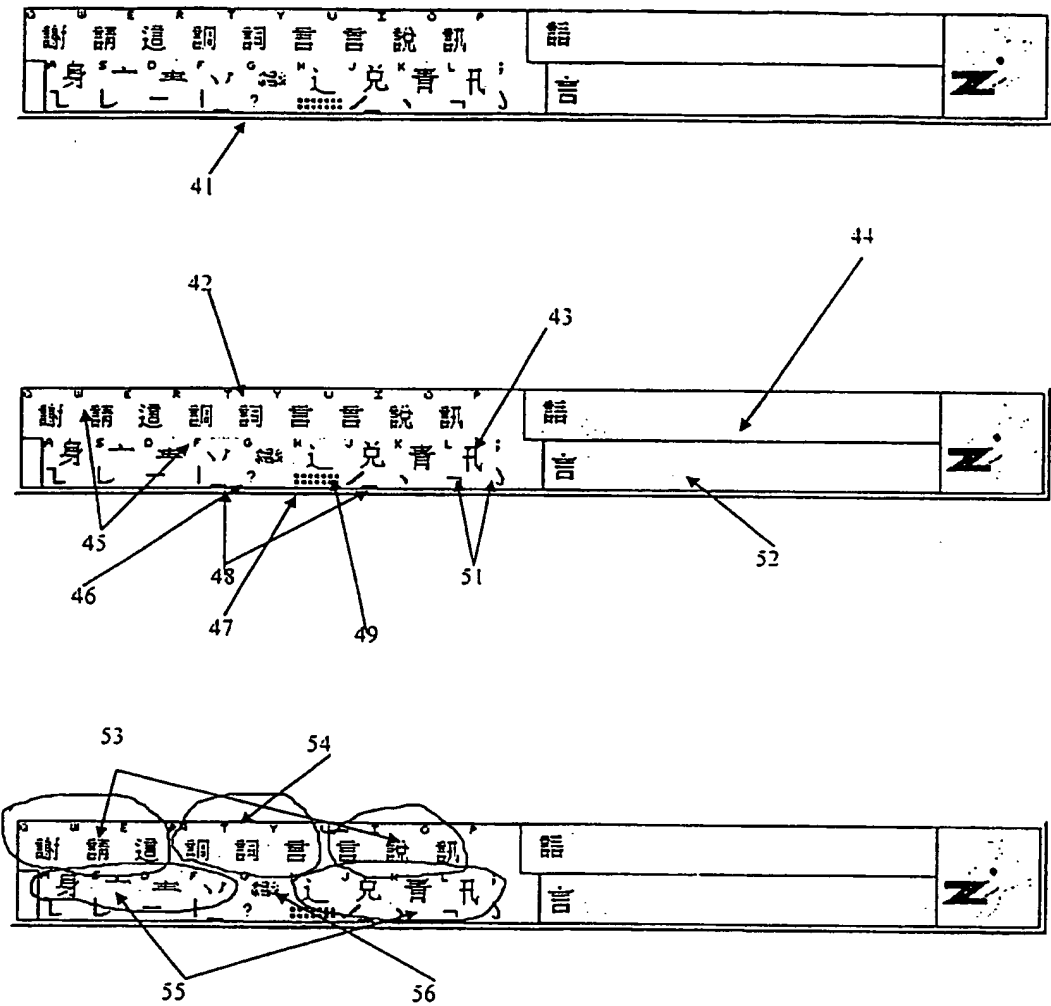


图 5

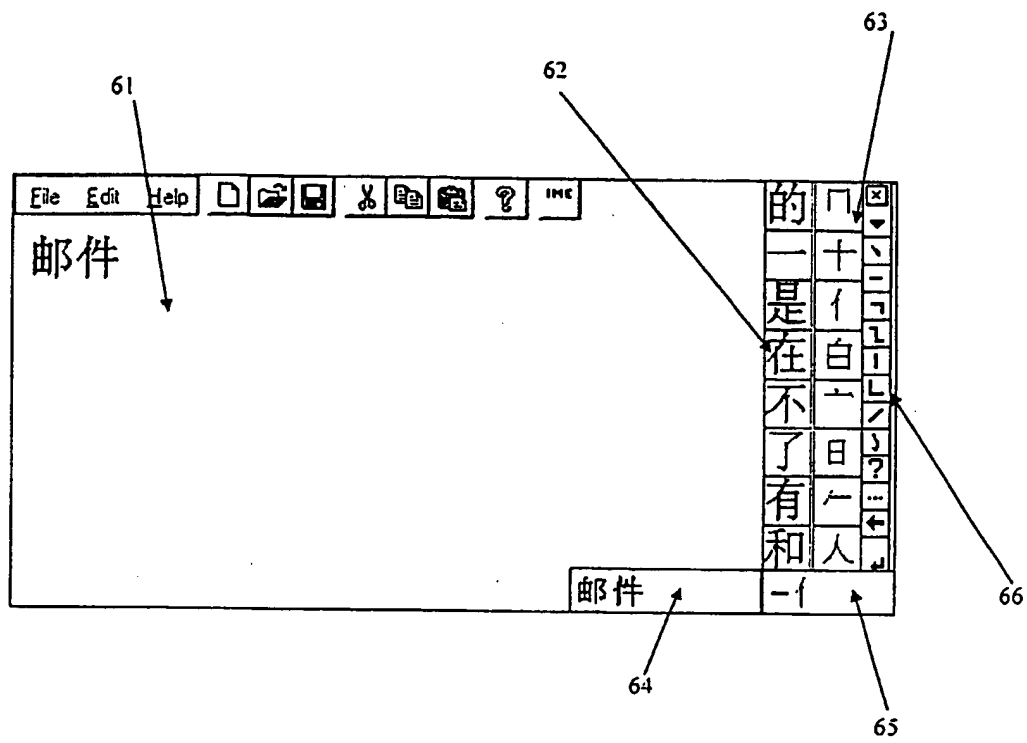


图 6

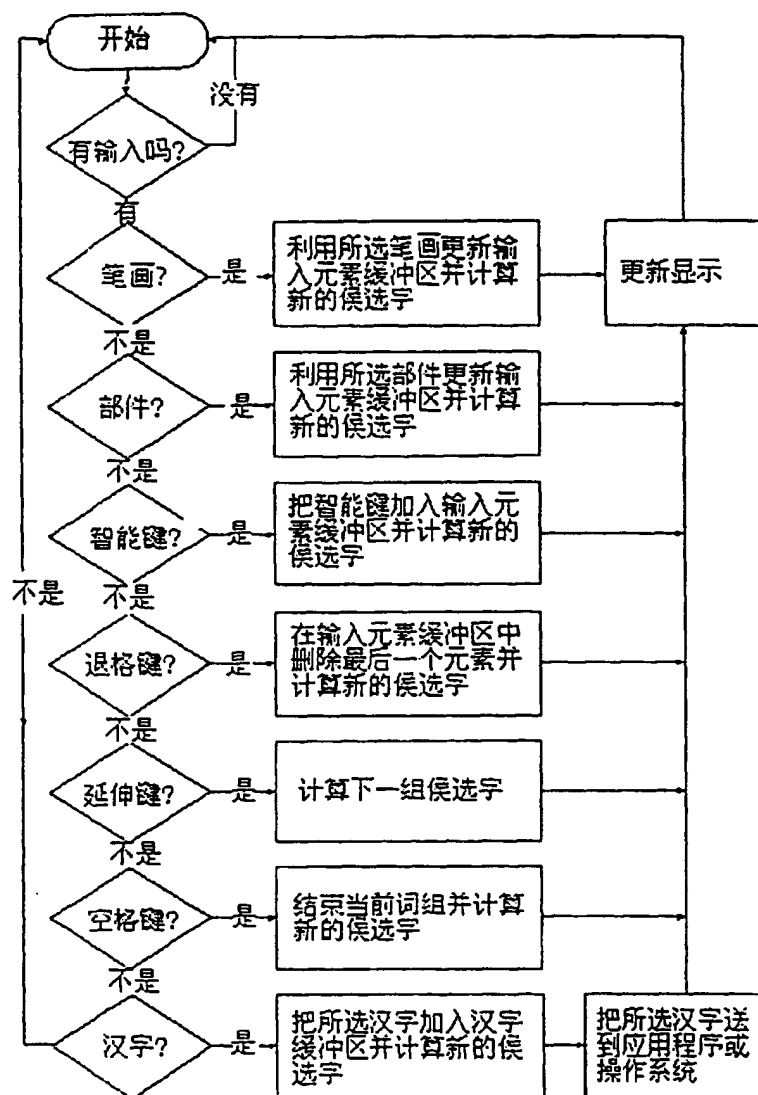


图 7

```

8.1    Character_Table: array of structure CharInfo

        Structure CharInfo:
8.2        UCHAR ucStrokes[12]; // Stroke composition Info
                                     // 24 strokes stored as nybbles
8.3        ULONG ulComponents: // 10 bit first drawn component
                                     // 10 bit second drawn component
                                     // 8 bit KangXi radical
                                     // 4 bits of properties
8.4        USHORT usFreq;        // frequency of the character
8.5        USHORT usCode;        // base codeset code of the character
8.6        wchar_t wsInitial[3]; // the initial sound of the character
        wchar_t wszFinal[7];    // the final sound of the character
8.7        ULONG ulOffsetWords; // pointer to words beginning with this char

8.8    Component Table: array of structure Component

        Structure Component:
8.9        UCHAR ucStrokes[12]; // Stroke composition Info
8.10       ULONG ulComponent;    // the first drawn component

8.11    Word Table: array of structure Word

        Structure Word:
        array of ULONG word[]; // code of the words
8.12       ULONG ulWordFreq;    // high bit is set + Word Freq.

8.13    Component Frequency Table:
        array of ULONG ulFreq; // size in: total number of components

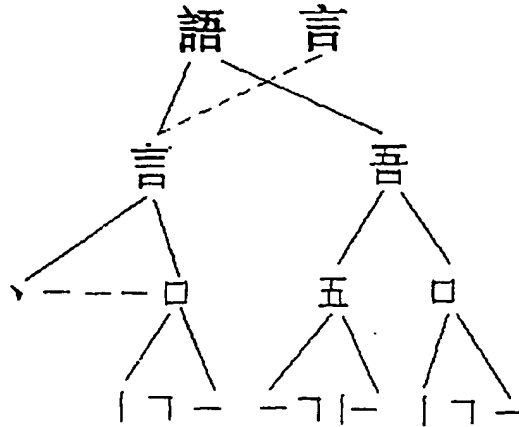
8.14    Complex Component Frequency Table:
        array of ULONG ulFreq; // size in: total number of components

8.15    Stroke Match Buffer: UCHAR ucStrokes[12];

8.16    Letter Buffer: array of char cLetters[];

```

图 8



开始:

人 大 一 不 的 是 我 有 在	繁體部件	
人 大 一 不 的 是 我 有 在		

第一步: Shift L

論 話 說 說 言 通 語 謝 該	繁體部件	
論 話 說 說 言 通 語 謝 該	言	

第二步: D

據 語 語 說 講 計 討 講 評	繁體部件	
據 語 語 說 講 計 討 講 評	言 -	

第三步: E

系 系 系 系 句 詞 云 文 將	語	
系 系 系 系 句 詞 云 文 將		

第四步: shift L

謝 語 通 說 詞 言 言 說 語	語	
謝 語 通 說 詞 言 言 說 語	言	

第五步: Y

在 有 我 是 學 的 不 一 大	語 言	
在 有 我 是 學 的 不 一 大		

图 9

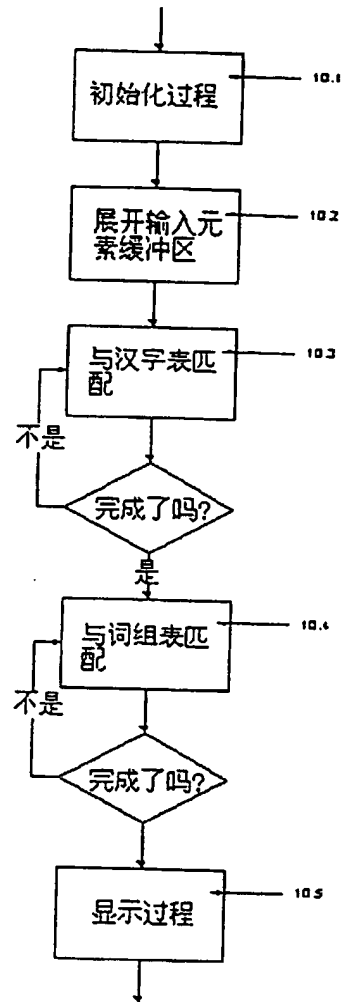


图 10

